Biomedical Optics EXPRESS

CCS-UNet: a cross-channel spatial attention model for accurate retinal vessel segmentation

YONG-FEI ZHU,[†] XIANG XU,[†] XUE-DIAN ZHANG, AND MIN-SHAN JIANG[*]

Shanghai Key Laboratory of Contemporary Optics System, College of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

[†]These authors contributed equally to this work.

[*]*jiangmsc@gmail.com*

Abstract: Precise segmentation of retinal vessels plays an important role in computer-assisted diagnosis. Deep learning models have been applied to retinal vessel segmentation, but the efficacy is limited by the significant scale variation of vascular structures and the intricate background of retinal images. This paper supposes a cross-channel spatial attention U-Net (CCS-UNet) for accurate retinal vessel segmentation. In comparison to other models based on U-Net, our model employes a ResNeSt block for the encoder-decoder architecture. The block has a multi-branch structure that enables the model to extract more diverse vascular features. It facilitates weight distribution across channels through the incorporation of soft attention, which effectively aggregates contextual information in vascular images. Furthermore, we suppose an attention mechanism within the skip connection. This mechanism serves to enhance feature integration across various layers, thereby mitigating the degradation of effective information. It helps acquire cross-channel information and enhance the localization of regions of interest, ultimately leading to improved recognition of vascular structures. In addition, the feature fusion module (FFM) module is used to provide semantic information for a more refined vascular segmentation map. We evaluated CCS-UNet based on five benchmark retinal image datasets, DRIVE, CHASEDB1, STARE, IOSTAR and HRF. Our proposed method exhibits superior segmentation efficacy compared to other state-of-the-art techniques with a global accuracy of 0.9617/0.9806/0.9766/0.9786/0.9834 and AUC of 0.9863/0.9894/0.9938/0.9902/0.9855 on DRIVE, CHASEDB1, STARE, IOSTAR and HRF respectively. Ablation studies are also performed to evaluate the the relative contributions of different architectural components. Our proposed model is potential for diagnostic aid of retinal diseases.

1. Introduction

Retinal vessel structure analysis has been extensively used to detect disorders like diabetic retinopathy [1,2]. The captured retinal vasculature can be used to evaluate the severity of retinal diseases and the efficacy of treatments. The reinal pathology is also related to other body abnormalities, such as diabetes, cirrhosis and nephritis. Thus, the segmentation of retinal vasculature is one of the crucial steps in ophthalmic clinic. The experienced doctors may focus on the morphological properties, such as length, width, curvature, or how the vessels branch and angle.

Accurate segmentation in retinal vasculature is of great assistance in aquring the morphological information. Nevertheless, manual vessel segmentation is tedious, time-consuming and heavily related to physician experience. Many minor, fragile, tightly connected vessels exist in the retina, and there is no apparent difference between the vessel part and the background. Furthermore, noise and uneven illumination might affect the fundus picture. At present, there are two main types of methods for retinal vessel segmentation: supervised learning methods and unsupervised learning methods.

#495766 https://doi.org/10.1364/BOE.495766
Journal © 2023 Received 18 May 2023; revised 14 Jul 2023; accepted 9 Aug 2023; published 18 Aug 2023

Unsupervised learning methods do not require the utilization of human-annotated labels as reference points. Unsupervised techniques can be classified into two distinct groups: matched filter-based approaches [3–5] and model-based approaches [6–8]. Mendonca et al. [9] employed four directional difference operators and morphological techniques to extract vascular centerlines. Segmented vascular trees were obtained by Fraz et al. [10] through the utilization of first-order derivatives of Gaussian filters in four directions and a multidirectional morphological top-hat operator. Zhao et al. [11] introduced an innovative infinite active contour model that leverages the blended region data from images to effectively segment blood vessels. Lam et al. [12] introduced a novel method for vessel segmentation using a multi-concave modeling approach. This approach incorporates three distinct metrics, namely the distinguishable concave metric, linear concave metric, and locally normalized concave metric. In general, unsupervised techniques do not necessitate training and labeling data, thereby significantly mitigating labor-intensive efforts. Unsupervised techniques typically need a feature extractor that is manually crafted, relying on prior knowledge and performing adequately for images with a solitary background. However, such methods are not satisfactory for learning retinal images with multiple features or more intricate backgrounds.

Supervised learning methods usually require expert-labeled samples to train the model, which can achieve relatively accurate results. Deep neural networks (DNNs) were introduced by Laskowski et al. [13] for vessel segmentation, and they showed good accuracy on the DRIVE dataset. DNNs and fully connected conditional random fields (CRFs) were merged by Fu et al. [14] who approached the vascular segmentation problem as a border detection challenge. In order to segment out more fine vessels and lessen false positives at vessel boundaries, Son et al. [15] suggested a strategy based on generative adversarial training. To cope with retinal images in complex backgrounds, including low-contrast vascular structures and lesion areas, Mo and Zhang [16] created a fully convolutional network based on deep supervision, which used multi-scale stratification features with different receptive field. Considering the high imbalance between the number of coarse and fine vessels, Yan et al. [17] proposed a three-stage retinal vessel segmentation network, namely coarse vessel segmentation, fine vessel segmentation and vessel fusion, and the learned discriminative features were able to better segment coarse and fine vessels. In addition, several studies have been devoted to improving the loss function, Yan et al. [18] examined a loss function that integrated segment-level and pixel-level losses in order to effectively account for the significance of vessels across various scales.

At present, U-Net [19] has recently gained popularity as the most effective architecture for vessel segmentation tasks and has demonstrated exceptional performance in the medical industry. In order to make up for the lost details caused by pooling operation, U-Net uses skip connection to fuse neighbouring hierarchical features, and produces impressive vessel segmentation results from a tiny dataset. The U-Net architecture's straightforward model structure and effective vascular segmentation capabilities have garnered increasing attention among scholars as a backbone network for fundus image segmentation. Consequently, numerous U-Net variants have been proposed. Lian et al. [20] introduced a novel approach for precise retinal vessel segmentation, which overcomes the limitations of other models that rely solely on global features and struggle to handle local intricacies. Their method involves an enhanced residual U-Net that incorporates both global and local features. DEU-Net [21] employed a spatial encoding path and a context encoding path to capture the intricate spatial and semantic details in vascular images. Additionally, it integrated a channel attention mechanism to facilitate feature map selection. CE-Net [22] proposed a contextual encoder network for the purpose of preserving spatial information in 2D medical image segmentation, while also capturing high-level semantic information. Li et al. [23] introduced a U-Net architecture that is lightweight and incorporates an attention mechanism within the decoder stage. This design aims to capture both global and augmented features. The attention gate module proposed by Oktay et al. [24] is designed to selectively suppress non-relevant factors in an image and concentrate on vessels of varying sizes, resulting in the production of precise vessel segmentation maps. In order to extract effective multiscale feature information and make full use of the deep feature map, NFN+ [25] used a cascaded U-Net structure to transfer the multiscale features and vascular probability map obtained from the shallow layer to the deep layer by skip connections, and the backend network further refines the map. Wu et al. [26] introduced an innovative scale and context-sensitive network (SCS-Net) to address the challenges posed by multiscale vascular changes and intricate vascular environments. The SCS-Net introduces a semantic aggregation module (SFA), which aims to facilitate the extraction of multi-scale information. Additionally, the network proposes an adaptive feature fusion module to improve the fusion of information across adjacent layers. Wang et al. [27] proposed a hard attention network, which mainly consists of three decoders, one decoder is used for the identification of hard and easy regions, and the other two are responsible for the segmentation of hard and easy regions, respectively. Some other techniques have been developed to improve the connection mechanism of U-Net, including augmenting the number of skip connections and employing multiple image coding paths to effectively capture information [28–30]. Some other approaches assist the network training by inserting filters in U-Net. Yin et al. [31] proposed a multi-scale input U-shaped network (SU-Net), which includes a guided image filter module to recover structural information through the guidance image. DF-Net [32] inserts a Frangi filter into the feature fusion module to obtain a compact yet domain invariant feature representation by fusing the vessel responses obtained from the filter with deep features. Not satisfied with the limitations of CNN on receptive field, some researches are devoted to integrate transformer into U-Net, and TransUNet [33] is better to extract features by adding transformer branch at the end of the encoder of U-Net, while UTNet [34] adds improved transformer module at the encoder and decoder stages respectively. This hybrid model uses the spatial induction bias specific to convolution to avoid large-scale pre-training on the one hand, and the transformer to capture global features on the other.

Despite the relatively favourable segmentation outcomes that the aforementioned U-shaped network can attain, it still exhibits deficiencies for retinal vessel segmentation. The limited feature extraction capability of many networks poses a challenge in extracting valuable vascular information from retinal images with low contrast, lesion area, optic disc, and optic cup in a complex semantic context. The vascular structure has multi-scale variations and unbalanced class distributions in retinal images. Therefore we need a more effective attention module to help the network to recognize the vascular structure adaptively. Considering the impact of downsampling operations, an effective fusion strategy is also crucial for aggregating spatial and semantic information.

Based on the above problems, we supposes a cross-channel spatial attention U-Net (CCS-UNet) for accurate retinal vessel segmentation. The primary components of the system are comprised of three fundamental modules. Firstly, we replace the original common convolutional block with ResNeSt block [35], which adopts a multi-branch structure to help the network identify more diverse vessel features. Split attention can help the network extract contextual information through channel attention. Secondly, considering the semantic differences between different feature layers, we introduce the CCS module into the skip connection for feature layer information fusion and accurate vessel structure recognition. Finally, deep supervision is introduced to provide the network with more semantic information through multiple lateral output layers to help the network training, while obtaining a more refined vascular segmentation map. In summary, this paper has three main contributions:

1) We supposed a novel network named CCS-UNet, which used ResNeSt as the backbone of the encoder-decoder architecture. Facing the retinal images under complex semantics, the multi-branch structure of the ResNeSt block helps the network extract more diverse vascular

features, and each subsequent grouping uses split attention for feature fusion, which effectively aggregates the contextual information around the vessels using the spatial attention mechanism.

2) We added the CCS module to the skip connection, which facilitates the integration of spatial and semantic information, compensates for the information loss caused by the downsampling operation. In addition, the resulting attention map helps the network obtain cross-channel information and improve the identification of regions of interest for better recognition of vascular structures.

3) The feature fusion module (FFM) was employed to enhance the segmentation accuracy by obtaining more comprehensive semantic representations from the side-output layers through the provision of supplementary supervision during the initial phases of the decoder network.

2. Methodology

The overall architecture of our designed CCS-UNet is shown in Fig. 1, which adopts a U-shaped structure design with five encoder levels and symmetrical decoder layers and mainly consists of three core modules, including ResNeSt block, CCS and FFM. The ResNeSt block is embedded in the encoder and is mainly used to enhance the feature extraction capability of the model, thus extract more diverse vascular features and aggregate contextual semantic information in retinal images under complex semantics. The CCS is used to guide the information fusion of adjacent feature layers. CCS recovers the information loss caused by the downsampling operation. The generated attention map aids the network in obtaining cross-channel information and improving the identification of regions of interest for better vascular structure detection. Finally, we inserted the FFM module to provide more semantic information of the network through multiple lateral outputs to obtain finer vascular segmentation maps.

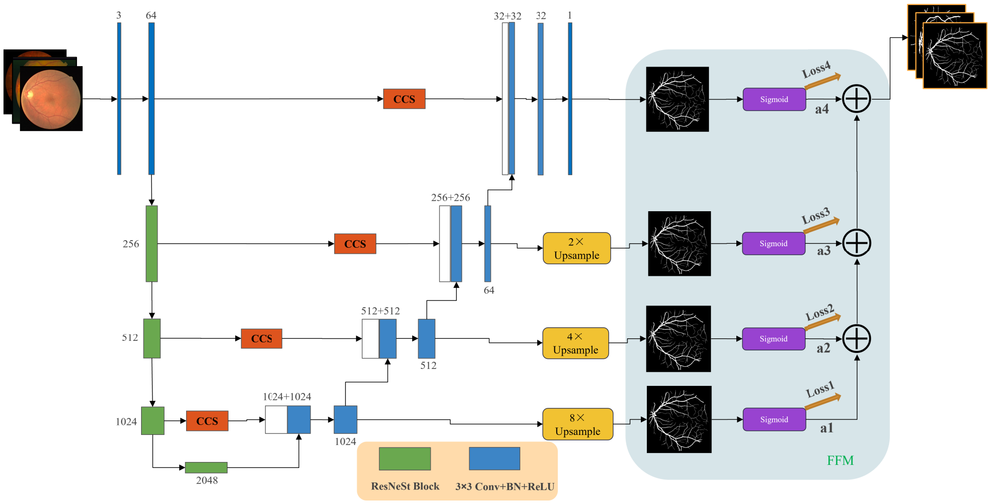


Fig. 1. The structure of our proposed CCS-UNet.

2.1. ResNeSt block

The ResNeSt [35] module is a residual architecture that incorporates a split attention mechanism. The primary characteristic of the ResNeSt module is its ability to consider a sequence of representations as a fusion of distinct feature groups, thereby providing focused attention to these groups. Fig. 2 depicts the intricate structure of ResNeSt. This work categorizes the input features $X \in R^{H\times W\times C}$ along the channel dimension into two equally cardinal groups. For every pair of

Biomedical Optics EXPRESS

cardinals, there are two parallel feature information streams. Each feature information branch contains a 1 × 1 convolutional layer, a 3 × 3 convolutional layer, Batch normalization, and ReLU layers, and the output feature map size is $H \times W \times C/4$.

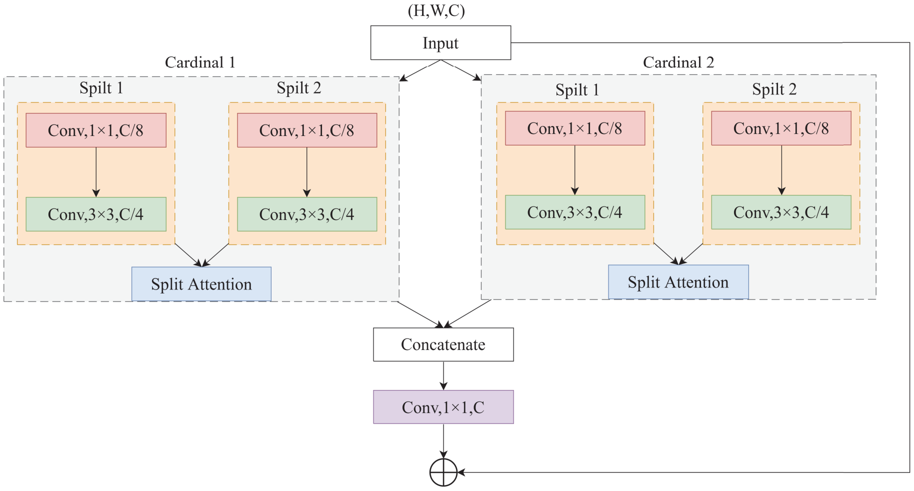


Fig. 2. The structure of the ResNeSt block.

In addition, to integrate the feature maps of each feature information branch, this paper adds a split attention module after each cardinal group. Fig. 3 depicts the split attention module. The initial step of the split attention module involves the fusion of feature maps from two distinct branches, which are identified as B_1 and B_2, via an element-wise summation operation. Channel-wise statistics S are computed through global pooling.

$$S = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [B_1(i,j) + B_2(i,j)] \tag{1}$$

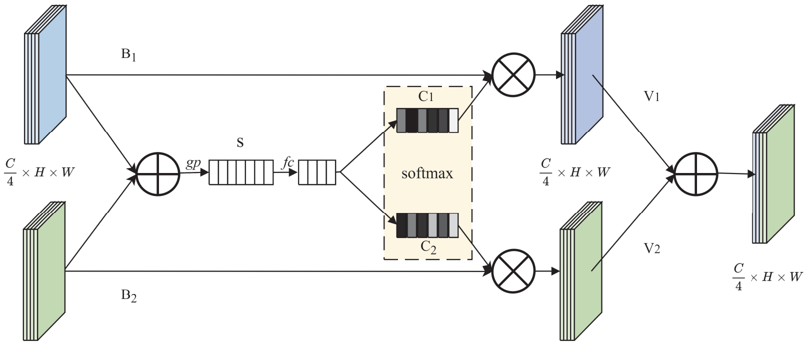

B1
B2
$\frac{C}{4} \times H \times W$
gp
S
fc
C1
softmax
C2
V1
V2
$\frac{C}{4} \times H \times W$
$\frac{C}{4} \times H \times W$

Fig. 3. The flowchart of the Split Attention.

S undergoes processing via two fully connected layers, followed by a softmax layer, resulting in the production of C_1 and C_2. C_1 and C_2 denote the soft attention weights assigned to the channel dimension of B_1 and B_2, respectively, which is utilized to produce:

$$\begin{cases} V_1 = C_1 \cdot B_1 \\ V_2 = C_2 \cdot B_2 \end{cases} \tag{2}$$

the symbol · denotes the element-wise multiplication across channels, the output of the cardinal group is

$$V = V_1 + V_2 \tag{3}$$

further, each cardinal group (denoted as U_1 and U_2) is merged across the channel axis, and then passes through a 1 × 1 convolution layer to generate:

$$Z = f^{1\times 1}([U_1, U_2]) \tag{4}$$

where $f^{1\times 1}(\cdot)$ represents the 1×1 convolution operation, $[\cdot, \cdot]$ is a symbol for a channel concatenation process. The final output of the ResNeSt block is generated through a shortcut connection:

$$Y = Z + T(X) \tag{5}$$

the function T denotes a suitable operation for transformation, such as stride, combined convolution with pooling, or residual connections with identity mapping to ensure the alignment of output shapes.

2.2. Feature fusion module

The task of retina vessel segmentation involves binary classification at the pixel level. This research employs a cross-entropy loss function to calculate the loss for individual lateral output layers. The ultimate loss function is derived by computing the average of the losses of these classifier layers. The inclusion of deep supervision during the training phase has been demonstrated to enhance the accuracy of the ultimate network [36]. Additionally, the prompt delivery of gradient information in the early stages of training is advantageous in mitigating gradient disappearance. The final loss function is formulated as:

$$L_{loss} = \frac{1}{N} \sum_{n=1}^{N} L_{cross-entropy}(y, y') \tag{6}$$

where N represents the number of layers of the side output layer, which is set to 4 in our study. $L_{cross-entropy}$ represents the cross-entropy loss, which is officially expressed as:

$$L_{cross-entropy}(y, y') = -\frac{1}{S} \sum_{i=1}^{S} (y_i \log(y_i{}') + (1 - y_i) \log(1 - y_i{}')) \tag{7}$$

where S represents the number of pixels of the retinal image, y' represents the predicted probability value, and y represents the true value of the label. As seen in Fig. 1, the network as a whole contains four classifier layers. Each decoder path is associated with a single output layer on one side, and the fifth classifier layer is formed by summing and averaging the preceding four layers. The ultimate forecast is regarded as the fifth layer of classification, which merges the four preceding classifiers.

2.3. Attention module

The CCS module was devised to achieve two objectives: firstly, to facilitate the efficient aggregation of spatial and semantic information, and secondly, to acquire cross-channel information and enhance the identification of regions of interest to improve the recognition of vascular structures. CCS module consists of two parallel branches, one used to obtain cross-channel information and the other to enhance the area of interest identification. As illustrated in Fig. 4, the input tensor $F \in R^{H\times W\times C}$ is initially received by the two branches of the attention module. The initial branch of the network encodes each channel of the input feature map F by applying two spatial domain pooling kernels, namely $(H, 1)$ and $(1, W)$, across both horizontally and vertically. The final coded feature maps Z^W and Z^H that aggregate features along the horizontal and vertical directions, respectively, are generated. This process aids in the precise identification of the region of interest. The mathematical representation of Z^H of the c-th channel at a given height h is expressed in the following manner:

$$Z_c^H(h) = \frac{1}{W} \sum_{0\leq i<W} X_c(h, i) \tag{8}$$

where $X_c(h, i)$ denotes the row vector of input features on a given channel c and height h. Similarly, the mathematical representation of Z^W of the c-th channel at a given width w is expressed in the following manner:

$$Z_c^W(w) = \frac{1}{H} \sum_{0\leq j<H} X_c(j, w) \tag{9}$$

where $X_c(j, w)$ denotes the column vector of input features on a given channel c and width w. The feature map acquired along different directions is concatenated in the spatial dimension. Subsequently, the 1×1 convolution layer is employed to generate the following:

$$F_{mid} = \delta(f^{1\times1}([Z^H, Z^W])) \tag{10}$$

where $[\cdot, \cdot]$ represents a concatenation operation along the spatial dimension, the non-linear activation function Swish is represented by the symbol δ. $f^{1\times1}$ denotes a convolutional operation with a convolutional kernel size of 1×1. The middle feature map, denoted by $F_{mid} \in R^{(H+W)\times(C/r)}$, is responsible for horizontal and vertical spatial information encoding. The variable r is utilized to regulate the magnitude of the channel dimension [37]. Then, the tensor F_{mid} was partitioned into two separate tensors $F^H \in R^{H\times(C/r)}$ and $F^W \in R^{W\times(C/r)}$ across the spatial dimension in the same way as originally concatenated. Ultimately, the channel dimensions of tensors F^H and F^W are adjusted to align with those of the input tensor F through the utilization of two 1×1 convolution layers.

$$G^H = \sigma(f^{1\times1}(F^H)) \tag{11}$$

$$G^W = \sigma(f^{1\times1}(F^W)) \tag{12}$$

where σ is the sigmoid function. Using the G^H and G^W as attention weights, generates:

$$F_C(i,j) = F(i,j) \cdot G^H(i) \cdot G^W(j) \tag{13}$$

where the symbol $\cdot$ denotes the element-wise multiplication across channels.

Similarly, in the second branch, first, the input feature map $F \in R^{H\times W\times C}$ are operated by averaging pooling and max pooling of the channel dimensions to generate feature maps ${F_S}^{MP} \in R^{H\times W\times1}$ and ${F_S}^{AP} \in R^{H\times W\times1}$ respectively. The feature map undergoes concatenation along the channel dimension, followed by the application of a 7×7 convolutional transform

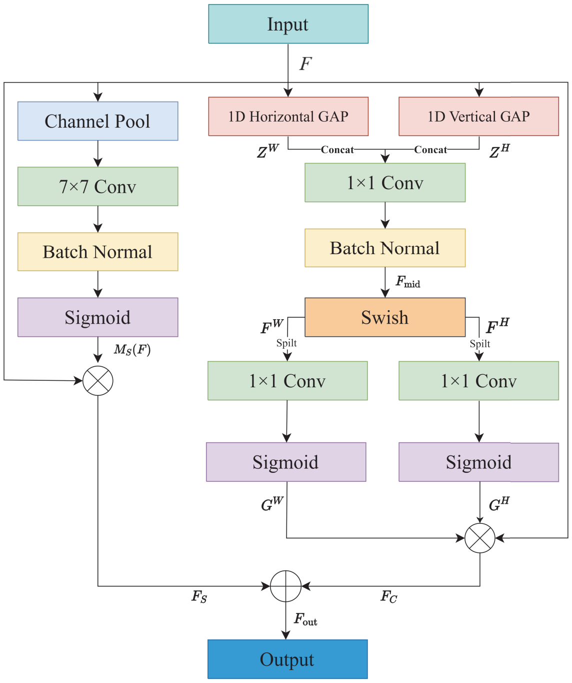


Fig. 4. The structure of cross-channel spatial attention.

function to produce a spatial attention map $M_S(F) \in R^{H\times W\times 1}$, finally generate:

$$\begin{aligned} F_S &= F \cdot M_S(F) \\ &= F \cdot \sigma\left(f^{7\times 7}([\mathrm{MaxPool}(F), \mathrm{AvgPool}(F)])\right) \\ &= F \cdot \sigma\left(f^{7\times 7}\left(\left[F_S^{MP}, F_S^{AP}\right]\right)\right) \end{aligned} \tag{14}$$

where $f^{7\times 7}$ denotes a convolutional operation with a convolutional kernel size of 7×7 and $\sigma(\mathbf{.})$ denotes the sigmoid activation function. $[\cdot, \cdot]$ represents a concatenation operation along the channel dimension. Finally, the output of the CCS module can be expressed as $F_{out} = (F_C + F_S)/2$ by averaging the two branches. Supplement 1 lists the definitions of the variables and the symbols.

3. Experiments

3.1. Data preparation

This paper employs five publicly accessible retinal vascular datasets: DRIVE [38], STARE [39], CHASEDB1 [40], IOSTAR [41], and HRF(High-Resolution Fundus) [42].

DRIVE: The DRIVE dataset was collected by a diabetic fundus screening organization from the Netherlands. There are a total of 40 fundus images in the DRIVE dataset, seven of which depict early diabetic retinopathy. Each image resolution is 565 × 584 pixels. We followed the official criteria and employed a sample of 20 images for the purpose of training, while the remaining 20 images are reserved for testing. The provided images exhibit the segmentation outcomes from two professionals along with their respective masks.

STARE: The STARE dataset comprises a total of 20 images, out of which 10 depict instances of retinopathy. Each image resolution is 605 × 700 pixels, and each image contains the segmentation results of two experts. 10-fold cross-validation was employed due to the absence of an official partition between training and test sets. We divided the dataset into 10 groups, and each group contained two images. In the training process, the training set consisted of 9 groups, while the test set comprised the remaining groups, with 10 training rounds until all images were tested.

CHASEDB1: The dataset denoted as CHASEDB1 is composed of a total of 28 fundus images, which were obtained from a sample of 14 children of school age. Each image resolution is of 999 × 960 pixels. This paper employed a methodology whereby the initial 20 images were designated for the purpose of training, while the remaining eight images were reserved for testing.

IOSTAR: The IOSTAR dataset consists of 30 scanninglLaser ophthalmoscopy (SLO) images. Each image resolution is 1024 × 1024 pixels. The dataset has been annotated by a group of experts in the domain of retinal image analysis, with each vessel being meticulously labeled. The initial 25 images were designated for the purposes of training while the remaining images were reserved for testing.

HRF: The HRF dataset comprises a total of 45 images, comprising 15 images for each category: healthy individuals, patients diagnosed with glaucoma, and patients diagnosed with diabetic retinopathy. The images are captured at a resolution of 3504 × 2336 pixels. The binary gold standard vascular segmentation image in each instance has been produced by a group of retinal image analysis scholars and physicians from associated eye hospitals. The initial 10 images from each category were chosen for the purpose of training, resulting in a training set of 30 images. In contrast, the remaining images were allocated to the test set.

3.2. Data augmentation

This study employed various data augmentation techniques to enhance the image generalization ability prior to network training. Specifically, a random horizontal flip with a probability of 0.5, a random rotation within the range of −20° to 20°, and a gamma contrast enhancement of 0.5 to 2 were utilized.

3.3. Implementation details

The PyTorch framework was utilized to train and validate the model suggested in this paper. We configured the network's learning rate to 0.01, and implemented a decay strategy whereby the learning rate was reduced to one-tenth of its current value in each 10 epoch. During the training phase, the model underwent training using the Adam [43] optimizer with default parameters while employing a binary cross-entropy loss function. We configured the training process to execute 100 epochs and terminated the training procedure upon reaching the maximum epoch value. The batch size was set to 2. To maintain experimental precision, all experimental procedures conducted in this paper were executed on a singular computer system. The GPU used in this computer is an NVIDIA GTX 3090–main software environment: python 3.6, Ubuntu 16.04, and PyTorch 1.7.0.

3.4. Evaluation criteria

This study employs nine frequently utilized evaluation metrics to assess the experimental outcomes, namely accuracy (ACC), sensitivity (SEN), specificity (SPE), false discovery rate (FDR), dice coefficient (DICE), gmean score [44], intersection over union (IOU), precision (PRE), and area under curve (AUC). These evaluation metrics can objectively show the strengths and weaknesses of different vessel segmentation methods. The SEN has the capacity to accurately represent the ratio of vessel pixels that are correctly identified. The SPE demonstrates the proportion of pixels correctly identified as non-vessel pixels. The evaluation criteria is defined as:

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \tag{15}$$

$$SEN = \frac{TP}{TP + FN} \tag{16}$$

$$SPE = \frac{TN}{TN + FP} \tag{17}$$

$$FDR = \frac{FP}{TP + FP} \tag{18}$$

$$DICE = \frac{2 \times TP}{FP + FN + 2 \times TP} \tag{19}$$

$$Gmean = \sqrt{SEN \times SPE} \tag{20}$$

$$IOU = \frac{TP}{FP + FN + TP} \tag{21}$$

$$PRE = \frac{TP}{FP + TP} \tag{22}$$

True positive (TP) refers to the vascular pixels that are accurately classified, while false positive (FP) pertains to the background pixels that are misclassified as the vascular. Likewise, TN (true negative) refers to pixels accurately categorized as background, while FN refers to pixels inaccurately classified as background. The model is also assessed by AUC, a higher value of AUC, approaching 1, indicates superior model performance.

4. Results

4.1. Comparisons with the state-of-the-art methods

In order to validate the superiority of the proposed CCS-UNet model, we conducted comparison experiments on five retinal image datasets. At the same time, eight prevalent methods, including U-Net [19], Att-UNet [24], U-Net++ [30], CE-Net [22], SU-Net [31], UTNet [34], TransUNet [33], and DF-Net [32], were selected to adopt the same training strategy and experimental environment.

4.1.1. Visual comparison

The visual performance of CCS-UNet is compared with other models in Fig. 5. In these models, U-Net has severe loss of edges for fine vessels. Att-UNet mainly provides an attention gate module to generate local weight maps for effective vessel recognition, but the simple structure limits the acquisition of global information. In contrast, although TransUNet and UTNet can effectively extract global information, their excessive model parameters are not suitable for retinal images with small datasets, and the overfitting problem still exists. SU-Net and DF-Net address the issue of information loss resulting from the downsampling procedure through the incorporation of filters, but their simple information fusion approach makes it challenging to effectively fuse different levels of features in the network, and many discontinuous vessels appear in their segmentation maps. Similarly, CE-Net employs multi-scale feature extraction modules to facilitate the extraction of semantic information, ignoring the role of feature fusion for the network and making it difficult to suppress noise effects.

In contrast, our proposed CCS-UNet obtains cross-channel information and improves the recognition of interest regions by introducing CCS in the skip connection. Our model effectively performs feature fusion to get a more continuous vessel segmentation map. Although U-Net++ achieves good results by fusing multiple feature layers of different scales, its feature extraction ability in the encoder stage must be improved for fine vessels segmentation. With the addition of the ResNeSt block, our CCS-UNet model is greatly enhanced in feature extraction capability and can cope with more complex semantic distribution of blood vessels. Fig. 6 shows the ROC curves for our CCS-UNet and eight other models across five datasets. Our ROC curve is closer to the upper left corner than others, which demonstrates that CCS-UNet has the higher accuracy of vessel segmentation in all of the five datasets.

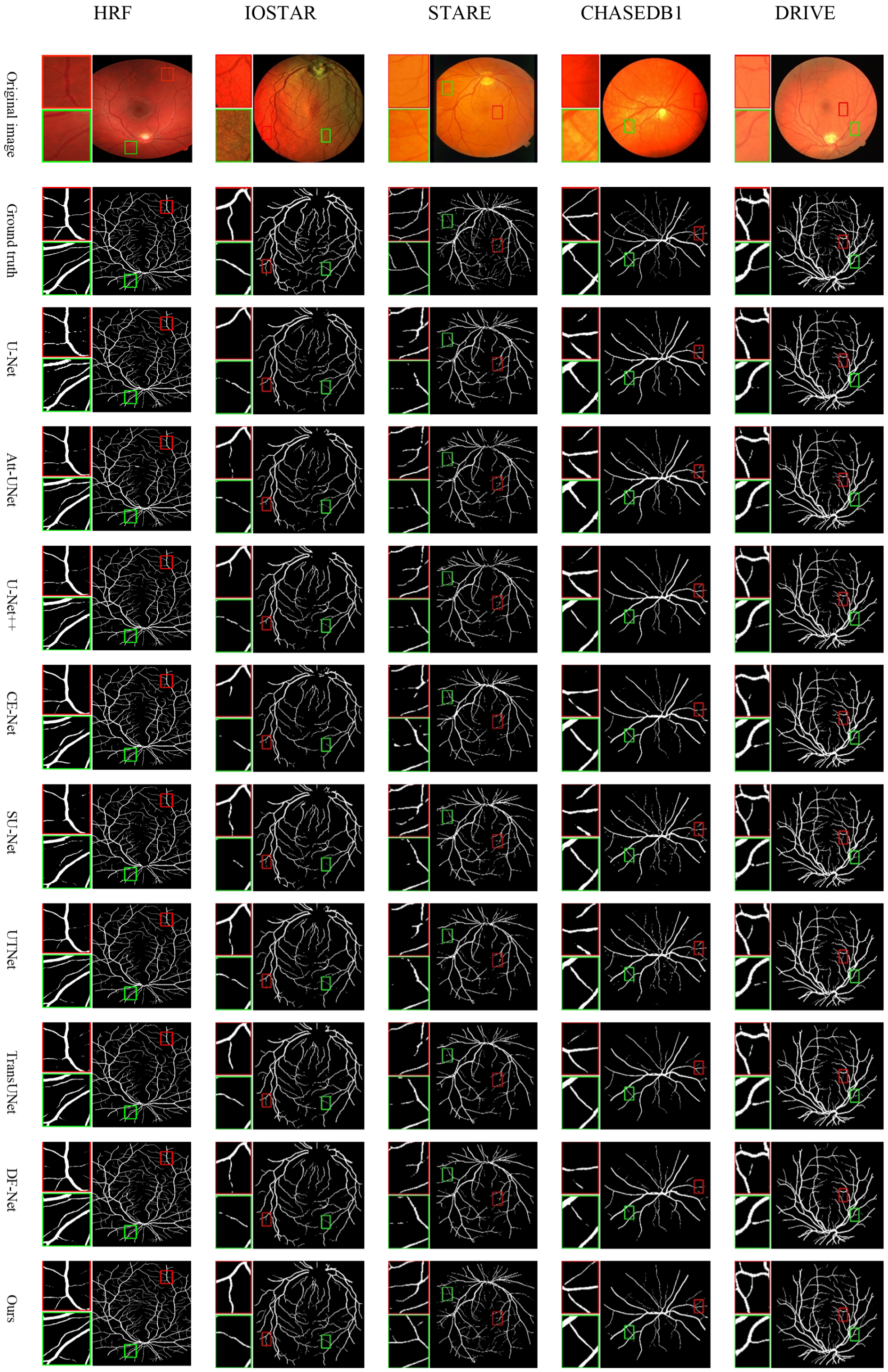
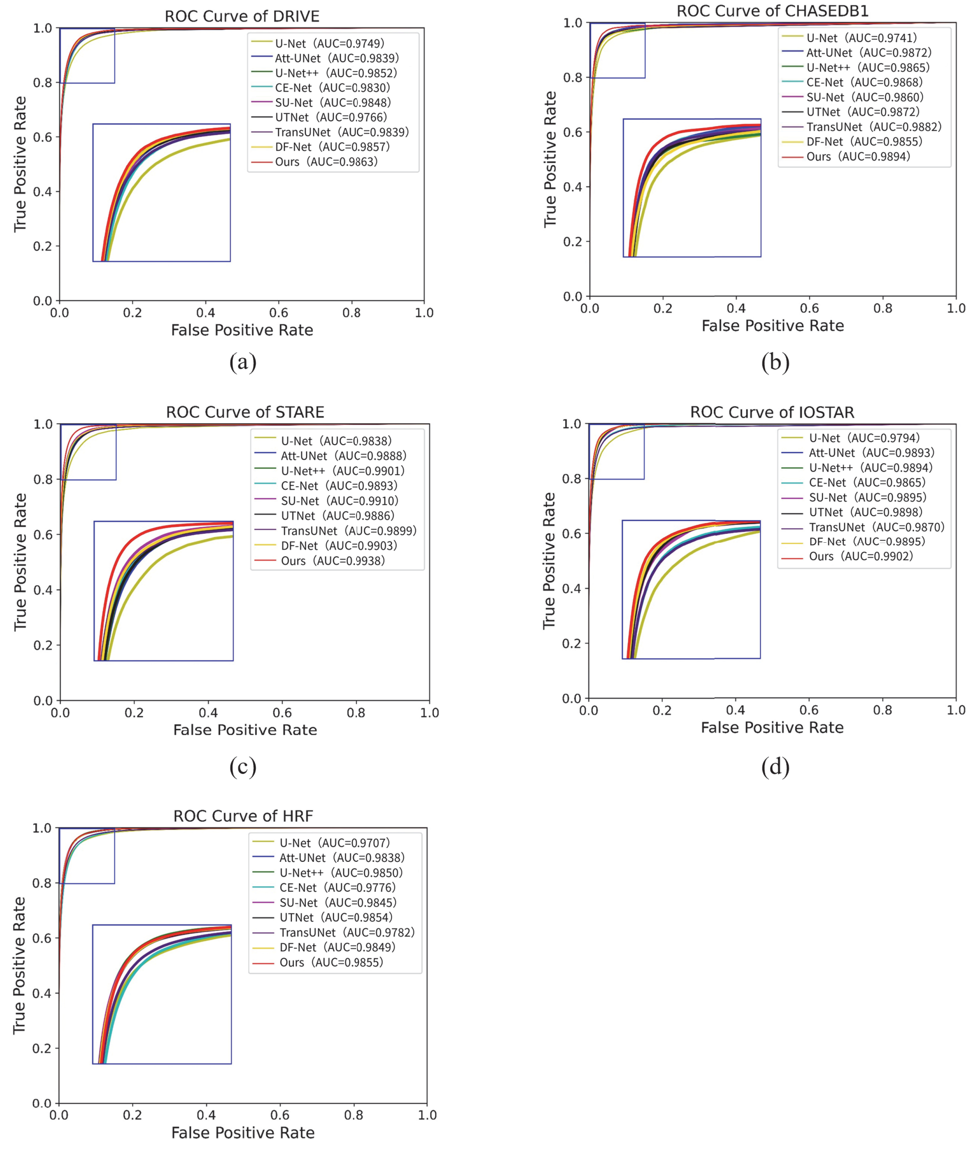


Fig. 5. Results of typical segmentation using different models in five classic datasets.



(a)

(b)

(c)

(d)

(e)

Fig. 6. ROC curves for different models: (a) DRIVE, (b) CHASEDB1, (c) STARE, (d) IOSTAR, (e) HRF.

4.1.2. Statistical evaluation

Quantitative methods is imperative for precisely and objectively evaluating experimental outcomes. We conducted comparison experiments on five datasets using nine evaluation metrics, and all models used the same experimental environment to guarantee fair comparison. According to the results presented in Table 1 for the DRIVE dataset, our proposed CCS-UNet model demonstrates superior performance across most evaluation metrics. Notably, the model achieves highest in AUC score 0.9863, ACC score 0.9617, and SEN score 0.7789, indicating that our network can identify more fine vessels. Although the SPE value (0.9839) was 0.17% lower compared to

Att-UNet (0.9856), our SEN value (0.7789) is 3.47% higher than Att-UNet (0.7442), and the other evaluation metrics of our model are also significantly superior. The results obtained for the CHASEDB1 dataset, as presented in Table 2, indicate that the CCS-UNet model attained the maximum values for all metrics except SPE. Compared with U-Net, the SEN value increases directly from 0.7553 to 0.8331, while the AUC and ACC values increase from 0.9741/0.9789 to 0.9894/0.9806, respectively. Only SPE is slightly lower than other methods, but the balance between our SEN and SPE is better. Table 3 records the experimental results for the STARE dataset, and again, our model still achieves the highest AUC, ACC, and SEN values compared to other models. In addition, we also conducted experiments on the HRF and IOSTAR datasets compared with other models, as presented in Table 4 and Table 5. We also presents an expanded set of evaluation metrics, namely IOU, gmean, FDR, PRE, and DICE, to comprehensively assess the model's performance on these five datasets. Table 1–5 demonstrate that CCS-UNet outperforms other methods in most metrics, indicating its superiority in vessel segmentation. The results demonstrate that our model exhibits superior vessel segmentation capabilities compared to other models when confronts with various retinal images containing intricate structures.

Table 1. Comparisons of existing approaches on DRIVE.

Method	Year	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net [19]	2015	0.9586	0.9749	0.7354	0.9856	0.6539	0.8505	0.0313	0.9686	0.9770
Att-UNet [24]	2018	0.9595	0.9839	0.7442	**0.9856**	0.6619	0.8556	0.0304	0.9596	0.9775
U-Net++[30]	2018	0.9602	0.9852	0.7553	0.9850	0.6685	0.8617	0.0291	0.9708	0.9778
CE-Net [22]	2019	0.9576	0.9830	0.7473	0.9830	0.6521	0.8566	0.0300	0.9699	0.9764
SU-Net [31]	2020	0.9598	0.9848	0.7538	0.9848	0.6663	0.8611	0.0293	0.9706	0.9776
UTNet [34]	2021	0.9601	0.9766	0.7617	0.9841	0.6700	0.8650	0.0283	0.9716	0.9778
TransUNet [33]	2021	0.9590	0.9839	0.7541	0.9838	0.6623	0.8608	0.0292	0.9707	0.9772
DF-Net [32]	2022	0.9608	0.9857	0.7704	0.9841	0.6771	0.8702	0.0284	0.9724	0.9782
Ours	2023	**0.9617**	**0.9863**	**0.7789**	0.9839	**0.6841**	**0.8748**	**0.0264**	**0.9735**	**0.9786**

4.2. Ablation studies

In order to assess the significance of individual modules within the proposed model, we incorporated each module into U-Net independently. Then the network was trained to utilize the DRIVE, CHASEDB1, STARE, IOSTAR, and HRF datasets, with the experimental design depicted in Fig. 7. The impact of each module on the performance of vascular segmentation in different datasets is presented in Table 6–10.

4.2.1. Effectiveness of the ResNeSt block

Initially, we examine the effectiveness of the ResNeSt [35] block. The suggested ResNeSt block, denoted as 'U-Net+ResNeSt block', as seen in Table 6, raises the AUC and SEN by 1.14% and 2.52% (from 0.9749/0.7354 to 0.9863/0.7606, respectively) and an increase of about 0.26% in ACC when compared to 'U-Net'. Besides that, all other indicators increased to some extent. To verify the reliability of the module, we also performed ablation experiments on four other retinal datasets. As seen in Table 7–10, most metrics on all datasets are improved with the addition of the ResNeSt block, which helps improve extract features and identifies more fine vessels.

4.2.2. Effectiveness of the CCS

The recommended CCS module is incorporated into the U-Net architecture, denoted as 'U-Net+CCS', and subsequently employed in the DRIVE dataset. As demonstrated in Table 6, compared to the U-Net, 'U-Net+CCS' enhances SEN and ACC performance from 0.7354/0.9586

Table 2. Comparisons of existing approaches on CHASEDB1.

Method	Year	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net [19]	2015	0.9789	0.9741	0.7553	0.9900	0.6230	0.8637	0.0120	0.9879	0.9889
Att-UNet [24]	2018	0.9796	0.9872	0.7986	0.9885	0.6429	0.8876	0.0098	0.9901	0.9893
U-Net++[30]	2018	0.9794	0.9865	0.8104	0.9876	0.6442	0.8942	0.0092	0.9907	0.9892
CE-Net [22]	2019	0.9786	0.9868	0.8280	0.9858	0.6400	0.9032	0.0083	0.9916	0.9887
SU-Net [31]	2020	0.9794	0.9860	0.7476	**0.9907**	0.6256	0.8601	0.0122	0.9877	0.9892
UTNet [34]	2021	0.9800	0.9872	0.7850	0.9894	0.6436	0.8808	0.0104	0.9895	0.9895
TransUNet [33]	2021	0.9803	0.9882	0.7730	0.9904	0.6433	0.8745	0.0110	0.9889	0.9896
DF-Net [32]	2022	0.9792	0.9855	0.7999	0.9886	0.6352	0.8884	0.0096	0.9903	0.9891
Ours	2023	**0.9806**	**0.9894**	**0.8331**	0.9878	**0.6637**	**0.9070**	**0.0081**	**0.9918**	**0.9898**

Table 3. Comparisons of existing approaches on STARE.

Method	Year	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net [19]	2015	0.9686	0.9838	0.7924	0.9842	0.6761	0.8829	0.0183	0.9816	0.9829
Att-UNet [24]	2018	0.9691	0.9888	0.7756	0.9862	0.6748	0.8742	0.0197	0.9802	0.9832
U-Net++[30]	2018	0.9702	0.9901	0.8183	0.9822	0.6846	0.8963	0.0161	0.9838	0.9830
CE-Net [22]	2019	0.9709	0.9893	0.7762	0.9881	0.6858	0.8752	0.0195	0.9804	0.9842
SU-Net [31]	2020	0.9726	0.9910	0.8004	0.9879	0.7059	0.8888	0.0176	0.9823	0.9851
UTNet [34]	2021	0.9704	0.9886	0.8233	0.9834	0.6968	0.8996	0.0156	0.9843	0.9839
TransUNet [33]	2021	0.9727	0.9899	0.7787	**0.9898**	0.6997	0.8774	0.0193	0.9806	0.9852
DF-Net [32]	2022	0.9719	0.9903	0.7757	0.9893	0.6938	0.8759	0.0197	0.9802	0.9847
Ours	2023	**0.9766**	**0.9938**	**0.8435**	0.9883	**0.7468**	**0.9129**	**0.0138**	**0.9861**	**0.9872**

Table 4. Comparisons of existing approaches on IOSTAR.

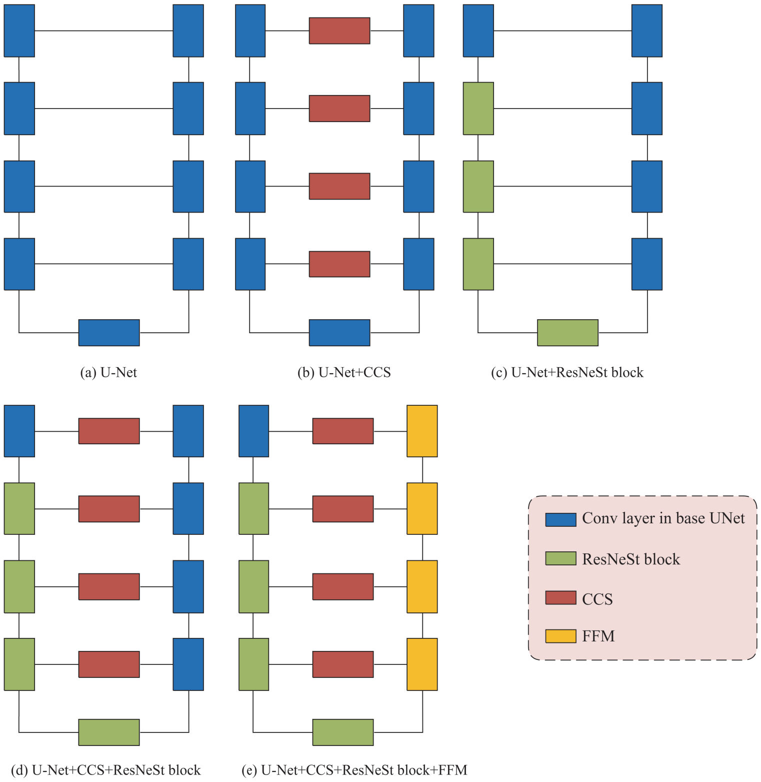
Method	Year	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net [19]	2015	0.9759	0.9794	0.8190	**0.9920**	0.7451	0.9001	0.0156	0.9843	0.9881
Att-UNet [24]	2018	0.9783	0.9893	0.8434	0.9903	0.7542	0.9128	0.0137	0.9862	0.9882
U-Net++[30]	2018	0.9782	0.9894	0.8491	0.9896	0.7543	0.9156	0.0131	0.9868	0.9882
CE-Net [22]	2019	0.9762	0.9865	0.8468	0.9874	0.7359	0.9136	0.0134	0.9865	0.9869
SU-Net [31]	2020	0.9781	0.9895	0.8769	0.9870	0.7594	0.9295	0.0109	0.9892	0.9881
UTNet [34]	2021	0.9773	0.9898	0.8601	0.9877	0.7500	0.9208	0.0123	0.9876	0.9876
TransUNet [33]	2021	0.9774	0.9870	0.8668	0.9871	0.7521	0.9242	0.0116	0.9883	0.9877
DF-Net [32]	2022	0.9783	0.9895	0.8727	0.9876	0.7600	0.9275	0.0111	0.9888	0.9882
Ours	2023	**0.9786**	**0.9902**	**0.8772**	0.9876	**0.7643**	**0.9301**	**0.0107**	**0.9892**	**0.9884**

to 0.7430/0.9589, respectively. Furthermore, as seen in Table 7–10, the addition of the CCS module also improves most metrics. The results indicate that with the addition of the CCS module, the network can perform more effective feature fusion and obtain more semantic information about the target vessels and a more continuous vessel segmentation map.

We also integrate both CCS and ResNeSt blocks into the U-Net, denoted as ’U-Net+CCS+ ResNeSt block’, in order to evaluate the potential synergistic effects of these two modules. Compared to the U-Net, the accuracy of segmentation has been significantly enhanced, as shown in Table 6, with an apparent increase of around 3.5% and 1.14% in terms of SEN and AUC, respectively, Even if there is a slight decrease in SPE (0.10%), this is acceptable. The results

Table 5. Comparisons of existing approaches on HRF.

Method	Year	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net [19]	2015	0.9809	0.9707	0.6983	0.9939	0.6194	0.8308	0.0138	0.9861	0.9900
Att-UNet [24]	2018	0.9823	0.9838	0.7024	0.9952	0.6369	0.8334	0.0136	0.9863	0.9907
U-Net++[30]	2018	0.9831	0.9850	0.7325	0.9949	0.6618	0.8523	0.0124	0.9875	0.9912
CE-Net [22]	2019	0.9812	0.9776	0.7146	0.9937	0.6325	0.8415	0.0132	0.9867	0.9902
SU-Net [31]	2020	0.9832	0.9845	0.7810	0.9927	0.6750	**0.8798**	**0.0103**	**0.9896**	0.9912
UTNet [34]	2021	0.9825	0.9854	0.7079	**0.9955**	0.6466	0.8382	0.0136	0.9863	0.9909
TransUNet [33]	2021	0.9810	0.9782	0.7248	0.9930	0.6344	0.8471	0.0128	0.9871	0.9871
DF-Net [32]	2022	0.9832	0.9849	0.7669	0.9932	0.6734	0.8722	0.0107	0.9892	0.9912
Ours	2023	**0.9834**	**0.9855**	**0.7671**	0.9936	**0.6769**	0.8720	0.0109	0.9890	**0.9913**



(a) U-Net (b) U-Net+CCS (c) U-Net+ResNeSt block

(d) U-Net+CCS+ResNeSt block (e) U-Net+CCS+ResNeSt block+FFM

Fig. 7. Different models examined in our ablation studies.

Table 6. Ablation studies of our proposed method on DRIVE.

Method	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net	0.9586	0.9749	0.7354	**0.9856**	0.6539	0.8505	0.0313	0.9686	0.9770
U-Net+CCS	0.9589	0.9762	0.7430	0.9850	0.6580	0.8548	0.0305	0.9694	0.9771
U-Net+ResNeSt block	0.9612	0.9863	0.7606	0.9856	0.6762	0.8651	0.0284	0.9715	0.9784
U-Net+CCS +ResNeSt block	0.9613	0.9863	0.7704	0.9846	0.6798	0.8702	0.0274	0.9725	0.9785
U-Net+CCS +ResNeSt block+FFM	**0.9617**	**0.9863**	**0.7789**	0.9839	**0.6841**	**0.8748**	**0.0264**	**0.9735**	**0.9786**

Table 7. Ablation studies of our proposed method on CHASEDB1.

Method	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net	0.9789	0.9741	0.7553	**0.9900**	0.6230	0.8637	0.0120	0.9879	0.9889
U-Net+CCS	0.9786	0.9757	0.7804	0.9883	0.6260	0.8775	0.0107	0.9892	0.9887
U-Net+ResNeSt block	0.9806	0.9870	0.7928	0.9900	0.6538	0.8853	0.0102	0.9897	0.9898
U-Net+CCS +ResNeSt block	0.9805	0.9889	0.8127	0.9888	0.6579	0.8962	0.0091	0.9908	0.9898
U-Net+CCS +ResNeSt block+FFM	**0.9807**	**0.9894**	**0.8331**	0.9878	**0.6637**	**0.9070**	**0.0081**	**0.9918**	**0.9898**

Table 8. Ablation studies of our proposed method on STARE.

Method	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net	0.9686	0.9838	0.7924	0.9842	0.6761	0.8829	0.0183	0.9816	0.9829
U-Net+CCS	0.9714	0.9863	0.8061	0.9860	0.6973	0.8912	0.0170	0.9829	0.9844
U-Net+ResNeSt block	0.9755	0.9936	0.8320	0.9883	0.7355	0.9066	0.0148	0.9851	0.9867
U-Net+CCS +ResNeSt block	0.9759	0.9937	0.8217	**0.9896**	0.7372	0.9017	0.0157	0.9842	0.9869
U-Net+CCS +ResNeSt block+FFM	**0.9766**	**0.9938**	**0.8435**	0.9883	**0.7468**	**0.9129**	**0.0138**	**0.9861**	**0.9872**

Table 9. Ablation studies of our proposed method on IOSTAR.

Method	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net	0.9759	0.9794	0.8190	**0.9920**	0.7451	0.9001	0.0156	0.9843	0.9881
U-Net+CCS	0.9770	0.9741	0.8457	0.9885	0.7435	0.9132	0.0133	0.9866	0.9875
U-Net+ResNeSt block	0.9785	0.9895	0.8667	0.9883	0.7602	0.9246	0.0116	0.9883	0.9883
U-Net+CCS +ResNeSt block	**0.9791**	0.9902	0.8615	0.9895	**0.7649**	0.9223	0.0121	0.9878	**0.9886**
U-Net+CCS +ResNeSt block+FFM	0.9786	**0.9902**	**0.8772**	0.9876	0.7643	**0.9301**	**0.0107**	**0.9892**	0.9884

Table 10. Ablation studies of our proposed method on HRF.

Method	ACC	AUC	SEN	SPE	IOU	Gmean	FDR	PRE	DICE
U-Net	0.9809	0.9707	0.6983	0.9939	0.6194	0.8308	0.0138	0.9861	0.9900
U-Net+CCS	0.9817	0.9728	0.7095	0.9944	0.6356	0.8383	0.0134	0.9865	0.9904
U-Net+ResNeSt block	**0.9838**	0.9850	0.7603	0.9943	**0.6791**	0.8683	0.0112	0.9887	**0.9915**
U-Net+CCS +ResNeSt block	0.9827	0.9848	0.7182	**0.9949**	0.6500	0.8436	0.0129	0.9870	0.9909
U-Net+CCS +ResNeSt block+FFM	0.9834	**0.9855**	**0.7671**	0.9936	0.6769	**0.8720**	**0.0109**	**0.9890**	0.9913

indicate the integration of CCS and ResNeSt block is beneficial, which is also confirmed in Table 7–10.

4.2.3. Effectiveness of the FFM

The FFM module is incorporated into the network architecture to enhance the semantic information by means of the lateral output layer, thereby leading to a more precise vascular segmentation map. According to Table 6, compared to 'U-Net+CCS+ResNeSt block', all metrics improve except SPE, which only decreases by 0.07%. Furthermore, to achieve a more precise assessment of the influence of the FFM module on the experimental outcomes, we performed ablation experiments on four other additional datasets. The results of these experiments further substantiate the efficacy of the FFM module.

4.3. Segmentation of disease images

Clinically, retinal lesions can easily lead to vessel segmentation discontinuity in the lesion area. To verify the adaptability of our CCS-UNet for lesion image segmentation, the results of four lesion images in the STARE dataset, including Diabetic lesion, Retinitis, Bleeding and Exudation are illustrated in Fig. 8. As shown in Fig. 8, both U-Net [19] and DF-Net [32] are not ideal for

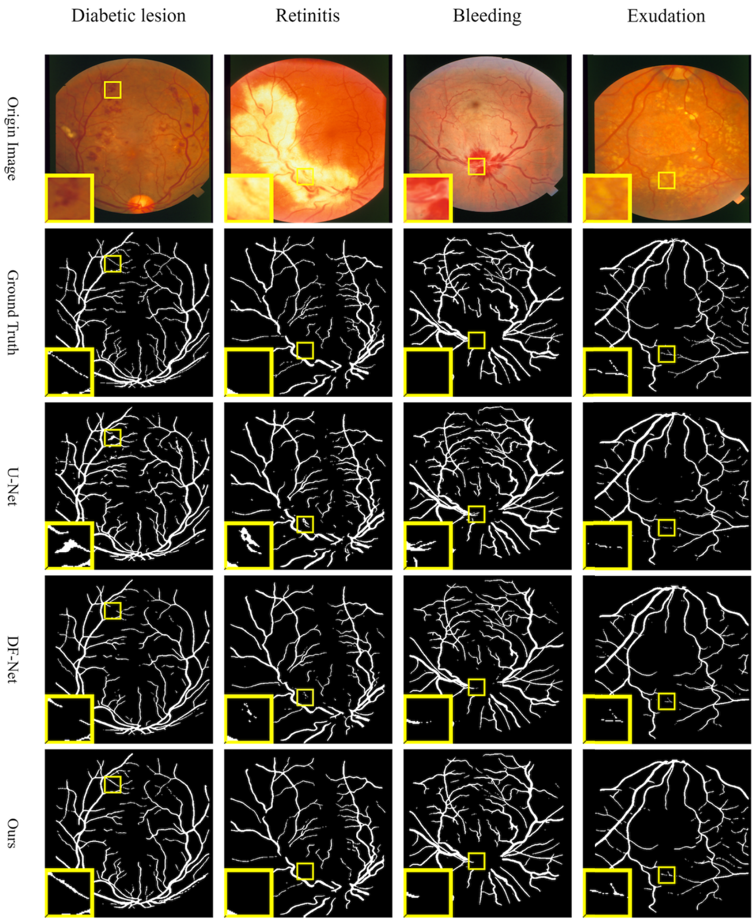


Fig. 8. Segmentation results of abnormal fundus images including Diabetic lesion, Retinitis, Bleeding and Exudation.

identifying lesion regions in the face of retinal images with complex backgrounds, however, our proposed CCS-UNet has strong anti-interference ability. (1) Diabetic lesion: Both U-Net and DF-Net show vascular discontinuity, while our CCS-UNet is able to segment a more continuous vessel. (2) Retinitis: Both U-Net and DF-Net show over-segmentation in the face of large lesions. Our CCS-UNet segmentation results for retinitis images are very similar to the ground truth, much better than U-Net and DF-Net. (3) Bledding: Although our CCS-UNet over-segments the spurious blood vessels, it also performs better than U-Net and DF-Net. (4) Exudation: Our CCS-UNet is able to restore the original detailed information more objectively, while U-Net and DF-Net are more lacking. In conclusion, with the well-designed attention module, our proposed CCS-UNet can distinguish the more detailed structure of the vessels from the complex vascular background.

4.4. Advantages and limitations

This study demonstrates that our suggested CCS-UNet outperforms other existing vascular segmentation techniques in terms of segmentation accuracy. As seen in Fig. 5, the model's feature extraction capability is improved through the incorporation of the ResNeSt block, which can extract more fine vessels in retinal images with complex backgrounds. In addition, inserting the CCS module in skip connection, an effective feature fusion approach, helps the network preserve cross-channel information while enhancing region of interest identification. Finally, introducing deep supervision provides the network with more semantic information, thus obtaining a complete vessel segmentation map. Although our model has the superior performance, it still has some limitations. Firstly, because of the introduction of ResNeSt Block, the model's parameter count and FLOPs have experienced a certain degree of augmentation. Secondly, owing to the overall scarcity of image quantity in every vessel dataset, the generalization ability of our model is limited to some extent. Even with the data augmentation strategy, the overfitting problem still exists, so more high-quality databases are needed. Finally, the segmentation performance of our model could be improved in the face of semantic contexts with more complex situations, such as images of lesion regions with ooze and low contrast. Better capture of contextual semantic information is a significant direction for model improvement.

5. Conclusion

This paper presents a CCS-UNet for increasing the accuracy of fundus vessel segmentation. CCS-UNet leverages ResNeSt as the encoder-decoder backbone to enhance the feature extraction performance, thereby facilitating the extraction of vascular information across various levels of complexity. We proposes a CCS module that can preserve cross-channel information while enhancing region of interest identification. Moreover, utilizing the FFM module aims to acquire a wider range of semantic information to improve vessel maps' quality. The proposed CCS-UNet is validated on five medical image datasets. The network proposed in this study exhibits superior performance across all five datasets compared to other contemporary techniques. In future research, we intend to employ this model in various medical imaging assignments, including but not limited to the aided diagnosis of pulmonary ailments.

Funding. National Natural Science Foundation of China (61905144); National Key Foundation for Exploring Scientific Instrument of China (2013YQ03065104).

Acknowledgments. This work was also supported by National Key Foundation for Exploring Scientific Instrument of China under Grant number 2013YQ03065104.

Disclosures. The authors declare no conflict of interest related to this article.

Data availability. The DRIVE dataset underlying the results presented in this paper are available in Ref. [38]. The STARE dataset underlying the results presented in this paper are available in Ref. [39]. The CHASEDB1 dataset underlying the results presented in this paper are available in Ref. [40]. The IOSTAR dataset underlying the results presented in this paper are available in Ref. [41]. The HRF dataset underlying the results presented in this paper are available in Ref. [42].

Supplemental document. See Supplement 1 for supporting content.

References

1. C. M. Oliveira, L. M. Cristovao, M. L. Ribeiro, and J. R. F. Abreu, "Improved automated screening of diabetic retinopathy," Ophthalmologica **226**(4), 191–197 (2011).
2. L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet, and J. P. Langlois, "Red lesion detection using dynamic shape features for diabetic retinopathy screening," IEEE Trans. Med. Imaging **35**(4), 1116–1126 (2016).
3. J. Elson, J. Precilla, P. Reshma, and N. S. Madhavaraja, Ieee, "Automated extraction and analysis of retinal blood vessels with multi scale matched filter," in *International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, (2017), pp. 775–779.
4. S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," IEEE Trans. Med. Imaging **8**(3), 263–269 (1989).
5. Y. Wang, G. Ji, P. Lin, and E. Trucco, "Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition," Pattern Recognit. **46**(8), 2117–2133 (2013).
6. L. Wang, A. Bhalerao, and R. Wilson, "Analysis of retinal vasculature using a multiresolution hermite model," IEEE Trans. Med. Imaging **26**(2), 137–152 (2007).
7. H. Narasimha-Iyer, J. M. Beach, B. Khoobehi, and B. Roysam, "Automatic identification of retinal arteries and veins from dual-wavelength images using structural and functional features," IEEE Trans. Biomed. Eng. **54**(8), 1427–1435 (2007).
8. H. Yan and B. S. Y. Lam, "A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields," IEEE Trans. Med. Imaging **27**(2), 237–246 (2008).
9. A. M. Mendona and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," IEEE Trans. Med. Imaging **25**(9), 1200–1213 (2006).
10. M. M. Fraz, S. A. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. R. Rudnicka, and C. G. Owen, "An approach to localize the retinal blood vessels using bit planes and centerline detection," Comput. Methods Programs Biomed. **108**(2), 600–616 (2012).
11. Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," IEEE Trans. Med. Imaging **34**(9), 1797–1807 (2015).
12. B. Lam, Y. Gao, and W. C. Liew, "General retinal vessel segmentation using regularization-based multiconcavity modeling," IEEE Trans. Med. Imaging **29**(7), 1369–1381 (2010).
13. P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016).
14. H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, (IEEE), pp. 698–701.
15. J. Son, S. J. Park, and K.-H. Jung, "Retinal vessel segmentation in fundoscopic images with generative adversarial networks," arXiv, arXiv:1706.09318 (2017).
16. J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," Int. J. CARS **12**(12), 2181–2193 (2017).
17. Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," IEEE J. Biomed. Health Inform. **23**(4), 1427–1436 (2019).
18. Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," IEEE Trans. Biomed. Eng. **65**(9), 1912–1923 (2018).
19. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer), pp. 234–241.
20. S. Lian, L. Li, G. Lian, X. Xiao, Z. Luo, and S. Li, "A global and local enhanced residual u-net for accurate retinal vessel segmentation," IEEE/ACM Trans. Comput. Biol. and Bioinf. **18**(3), 852–862 (2021).
21. B. Wang, S. Qiu, and H. He, "Dual encoding u-net for retinal vessel segmentation," in *International conference on medical image computing and computer-assisted intervention*, (Springer), pp. 84–92.
22. Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," IEEE Trans. Med. Imaging **38**(10), 2281–2292 (2019).
23. X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," IEEE Trans. Ind. Inf. **17**(3), 1958–1967 (2021).
24. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, and B. Kainz, "Attention u-net: Learning where to look for the pancreas," arXiv, arXiv:1804.03999 (2018).
25. Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Nfn plus : A novel network followed network for retinal vessel segmentation," Neural Networks **126**, 153–162 (2020).
26. H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," Med. Image Anal. **70**, 102025 (2021).

27. D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," IEEE J. Biomed. Health Inform. **24**(12), 3384–3396 (2020).
28. S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "Ccnet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features," Neurocomputing **392**, 268–276 (2020).
29. L. Li, M. Verma, Y. Nakashima, H. Nagahara, R. Kawasaki, and I. C. Soc, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, (2020), IEEE Winter Conference on Applications of Computer Vision, pp. 3645–3654.
30. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, *Unet++: A nested u-net architecture for medical image segmentation* (Springer, 2018), pp. 3–11.
31. P. Yin, R. Yuan, Y. Cheng, and Q. Wu, "Deep guidance network for biomedical image segmentation," IEEE Access **8**, 116106 (2020).
32. P. Yin, H. Cai, and Q. Wu, "Df-net: Deep fusion network for multi-source vessel segmentation," Inf. Fusion **78**, 199–208 (2022).
33. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv, arXiv:2102.04306 (2021).
34. Y. Gao, M. Zhou, and D. N. Metaxas, "Utnet: a hybrid transformer architecture for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer), pp. 61–71.
35. H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. Smola, Ieee, "Resnest: Split-attention networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), pp. 2735–2745.
36. W. Ke, J. Chen, J. Jiao, G. Zhao, and Q. Ye, "Srn: Side-output residual network for object symmetry detection in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1068–1076.
37. J. Hu, L. Shen, and G. Sun, Ieee, "Squeeze-and-excitation networks," in *31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2018), IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.
38. J. Staal, M. D. Abrámoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Trans. Med. Imaging **23**(4), 501–509 (2004).
39. A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Trans. Med. Imaging **19**(3), 203–210 (2000).
40. C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program," Invest. Ophthalmol. Vis. Sci. **50**(5), 2004–2010 (2009).
41. J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," IEEE Trans. Med. Imaging **35**(12), 2631–2644 (2016).
42. J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, and E. Angelopoulou, "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database," IET Image Processing **7**, 373–383 (2013).
43. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv, arXiv:1412.6980 (2014).
44. Z. Fan, J. Lu, C. Wei, H. Huang, X. Cai, and X. Chen, "A hierarchical image matting model for blood vessel segmentation in fundus images," IEEE Trans. on Image Process. **28**(5), 2367–2377 (2019).